

# **EXHIBIT G**

# Blobworld: A System for Region-Based Image Indexing and Retrieval (long version)\*

Chad Carson, Megan Thomas, Serge Belongie,  
Joseph M. Hellerstein, and Jitendra Malik

EECS Department, University of California, Berkeley, CA 94720, USA  
{carson,mct,sjb,jmh,malik}@eecs.berkeley.edu

**Abstract.** Blobworld is a system for image retrieval based on finding coherent image regions which roughly correspond to objects. Each image is automatically segmented into regions (“blobs”) with associated color and texture descriptors. Querying is based on the attributes of one or two regions of interest, rather than a description of the entire image. In order to make large-scale retrieval feasible, we index the blob descriptions using a tree. Because indexing in the high-dimensional feature space is computationally prohibitive, we use a lower-rank approximation to the high-dimensional distance. Experiments show good results for both querying and indexing.

## 1 Introduction

From a user’s point of view, the performance of an information retrieval system can be measured by the quality and speed with which it answers the user’s information need. Several factors contribute to overall performance:

- the time required to run each individual query,
- the quality (precision/recall) of each individual query’s results, and
- the understandability of results and ease of refining the query.

All of these factors should be considered together when designing a system. In addition, image database users generally want to find images based on the *objects* they contain, not just low-level features such as color and texture [5, 7]; image retrieval systems should be evaluated based on their performance at this task.

Current image retrieval systems tend to perform queries quickly but do not succeed in the other two areas. A key reason for the poor quality of query results is that the systems do not look for meaningful image regions corresponding to objects. Additionally, the results are often difficult to understand because the system acts like a black box. Consequently, the process of refining the query may be frustrating. When individual query results are unpredictable, it is difficult to produce a stream of queries that satisfies the user’s need.

---

\* A shorter version of this paper will appear at the International Conference on Visual Information Systems ’99 and is © Springer-Verlag.